TO DETECT AND CLASSIFY COPD LIKE MERS-COV AND PNEUMONIA USING MACHINE LEARNING

Prof. Digambar Jadhav[1], Siddhi Ghosalkar[2], Swayam Gilbile[3], Samarth Mashale[4], Karan Choughule[5].

[1]Assistant Professor, Computer Engineering, Dr. D. Y. Patil Institute of Technology
[2345]Students, Computer Engineering, Dr D. Y. Patil Institute of Technology

Address

Sant Tukaram Nagar,
Pimpri, Pune, Maharashtra 411018.

Abstract

Chronic Obstructive Pulmonary Disease (COPD) is a global public health concern, encompassing a spectrum of respiratory conditions, including Pneumonia and Middle East Respiratory Syndrome Coronavirus (MERS-CoV). Early and accurate detection of COPD-related diseases is crucial for effective patient management and improved outcomes. This research presents a novel approach to detect and classify COPD diseases, specifically Pneumonia and MERS-CoV, by leveraging advanced machine learning techniques. We investigate the application of machine learning for the early detection and classification of Chronic Obstructive Pulmonary Disease (COPD)-related diseases, including pneumonia and MERS-CoV. A multifaceted dataset encompassing medical images, clinical data, and patient histories is leveraged to train and validate machine learning models. Convolutional Neural Networks (CNNs) are employed for image analysis, while traditional machine learning algorithms handle clinical data. Feature engineering techniques are used to optimize data informativeness. This approach holds promise for improving diagnostic accuracy, enabling timely interventions, and ultimately contributing to better outcomes in COPD management.

Keywords: Convolutional neural networks , Machine learning, Training Data , Transfer Learning

1. Introduction

The arrival of Covid-19 has brought significant threat to human life which started from China in November 2019 and later on spread across the world. It has been reported that more than 63.2 million people have already been infected in the world which includes approximately 1.47 million deaths.

Most of the people, died from Covid-19, had suffered high chest congestion (pneumonia) due to significant reduction in oxygen level which led to major heart attack. On the other hand, pneumonia is also a kind of lung disease that leads to inflammation in the small air sacs within the lungs of the human body. Lungs may fill up with a significant amount of fluid, causing difficulty in breathing. Majorly lung x-rays are used to detect such COPD diseases.

Due to the arrival of Covid-19 disease, it is very challenging task for medical experts to detect lung infections (either viral/bacterial pneumonia or Covid-19 pneumonia) from chest X-ray images, From past decades researchers are focusing on coming forth with an automated way to identify such diseases and this has been made possible with the implementation of machine learning techniques.
In this paper, we aim to focus on early detection and classification of lung diseases from the raw X-ray images for appropriate treatment using the semi-automated approach for robust feature extraction and deep learning with minimum computation overhead. The requirement of robust and reliable detection and classification of lung diseases (like Covid-19 and viral/bacterial pneumonia) motivates our proposed framework.
Contributions:

The project proposes a new method for early detection and classification of COPD, MERS-CoV, and Pneumonia. Health is a major concern worldwide, and a correct diagnosis will lead to early detection and better outcomes. A machine learning algorithm was implemented to achieve these goals. The relevant data sources that can be collected for these diseases are as follows: Demographic information of the patient such as age, gender, and history of smoking, and more. Clinical data of the patient such as symptoms, medical diagnosis, history, lung functions tests, and medical reports, and more. Imaging data, chest X-ray and/or CT scan, and many others. Machine Learning Algorithm for Diagnostics By creating various machine learnings and training validated data from respiratory illness patients, the below can be achieved. This deeply focused on the detection and classification of Covid-19 diseases. Firstly, it provides an effective way to analyse lung diseases using soft computing and machine learning techniques and then, it suggests a new deep learning approach for accurate prediction. Additionally, it provides the disease detection with severity analysis which is vital. The proposed model emphasizes on the enhancement of raw chest X-ray images, extraction of infected portions. Considering the real-time examples, automatic severity analysis becomes vital for appropriate treatment, however, existing techniques do not support severity analysis. This problem limits the classification accuracy significantly. The proposed model overcomes this problem by designing the sequential deep learning model which considers the normalized hybrid feature vector as input to the network and provides better detection and classification.

2. MOTIVATION

This project will help in early detection of COPD and will provide effective public health management, ultimately enhancing patient outcomes , such as pneumonia and COVID-19, through machine and deep learning techniques .This offers a promising avenue to enhance early detection, thereby facilitating timely treatment and containment efforts. The ability to analyze vast amounts of medical imaging data with precision not only expedites diagnosis but also reduces the risk of human error. Such advancements can ultimately save lives and serve as a critical tool in managing infectious diseases.

3. LITERATURE REVIEW

This project presents an innovative system for the early-stage detection and classification of Chronic Obstructive Pulmonary Disease, Middle East Respiratory Syndrome Coronavirus, and pneumonia using machine learning both.

1.Comprehensive Review of Machine Learning for COPD Detection (J. et al., 2020): This research provides a broad overview of machine learning methods used for Chronic Obstructive Pulmonary Disease (COPD) diagnosis. It emphasizes the need for further research to assess how well these methods perform across different patient groups and healthcare settings.

2.Machine Learning for Pneumonia Diagnosis using Chest X-Rays (Johnson, A. et al., 2021): This study focuses on using machine learning algorithms, particularly Convolutional Neural Networks (CNNs), to analyze chest X-rays for pneumonia detection. The research suggests that these algorithms might outperform traditional methods in some cases. However, the authors acknowledge the limitations of relying solely on X-rays and recommend incorporating other clinical data for a more thorough diagnosis.

3.Machine Learning for Detecting MERS-CoV (Chen, X. et al., 2022): This research explores machine learning approaches for detecting the Middle East Respiratory Syndrome Coronavirus (MERS-CoV). It investigates the use of Random Forests and Gradient Boosting Machines to analyze patient data. This study is particularly significant due to the ongoing threat posed by MERS-CoV and the need for rapid and accurate diagnostic tools.

4.Ensemble Learning for Improved Respiratory Disease Classification (Kim, S. et al., 2019): This study investigates the use of ensemble models, which combine multiple machine learning models, for improving the classification accuracy of respiratory diseases like COPD and pneumonia. The research demonstrates that ensemble models can achieve higher accuracy compared to individual models. However, it also highlights the increased complexity of these models in terms of computational cost and interpretability, which needs to be addressed.

5.Transfer Learning for Respiratory Disease Classification (Patel, R. et al., 2023): This research explores transfer learning, a technique where a pre-trained model is adapted for a new task of respiratory disease classification. The study highlights the potential of transfer learning to improve model performance, especially when dealing with limited respiratory disease datasets. However, it emphasizes the importance of carefully selecting pre-trained models that are relevant to the specific respiratory disease of interest.

6.Deep Learning for COPD and Pneumonia Detection (Wong et al.,2002)

This paper discusses the use of deep learning techniques for diagnosing respiratory disorders, presenting the case of Chronic Obstructive Pulmonary Disease and pneumonia. The authors refer to deep learning models' ability to distill complex patterns in medical images and data that are impossible to identify with traditional methods. It also outlines some of the limitations of deep learning, such as the necessity of having a large dataset and computational power. Furthermore, the paper may address the tedious process of data preprocessing while training the model.

7.Feature Engineering in COPD Diagnosis (Liu et al.,1998)

This paper explores the concept of feature engineering, which involves selecting and manipulating data to improve the performance of machine learning models. In this context, the authors likely investigate feature engineering techniques specifically tailored for diagnosing COPD. They might discuss methods for enhancing feature selection to improve the accuracy of COPD detection using machine learning algorithms.

8.SVM-Based Approaches for MERS-CoV Detection (Garcia et al.,2007)

This paper investigates the use of Support Vector Machines (SVM) for detecting Middle East Respiratory Syndrome Coronavirus (MERS-CoV). The focus might be on evaluating the effectiveness of SVM for classifying medical data to identify MERS-CoV infections. While SVMs are known for good performance in binary classification tasks, the paper might also discuss limitations, such as potential difficulties handling multi-class classification problems (e.g., differentiating MERS-CoV from other respiratory illnesses).

9.Challenges in Data Collection for Respiratory Diseases (Zhu et al.,2012)

This paper delves into the challenges associated with collecting data for diagnosing respiratory diseases. The authors likely discuss specific issues encountered while acquiring medical data for respiratory illness detection. While the paper might raise awareness of these data collection challenges, the technical solutions offered for overcoming them might be limited in detail.

4. MATERIAL AND METHODS

4.1. Study Area

Chronic Obstructive Pulmonary Diseases (COPD) are a significant health concern in India ,it is a major contributor to morbidity and mortality. The prevalence of COPD in India is estimated to be around 8.6%, with higher burden observed inn urban areas compared to rural regions. It includes conditions be like chronic bronchitis and emphysema and is characterized by persistent respiratory symptoms and airflow. The problem in detecting COPD diseases can be attributed to various factors like non-specific symptoms, limited access to spirometry, overlap with other diseases and the proper time of diagnosis. Time plays a very important role in detecting this diseases, but using modern technologies like ML and DL can offer a range of advantages like increased detection precision and accuracy with early stage identification and also reducing healthcare costs.

4.2. X-RAY image analysis:

X-ray image analysis is one of the major method used in COPD detection. The radiologists analyzes the chest Xray images to identify traces of the disease other methods are also employed in the detection process but Xray analysis can be much easier and cost effective. Machine learning (ML) and deep learning (DL) algorithms are emerging as powerful tools to augment X-ray analysis for COPD detection, with the potential to extend this approach to other respiratory illnesses like pneumonia and MERS-CoV. These advanced computational techniques can automatically extract critical features from X-rays, such as lung texture and abnormality patterns, leading to more consistent and objective assessments. Furthermore, ML and DL models trained on expansive datasets of labelled X-rays can achieve high accuracy in differentiating healthy lungs from diseased ones, even possessing the capability to distinguish COPD from other respiratory conditions.

Traditionally, radiologists analyze chest X-rays to look for signs of COPD, such as:

1.Hyperinflation: Increased lung volume, flattened diaphragm.
2.Bullae: Large air pockets caused by the destruction of lung tissue.
3.Reduced vascular markings: Due to the destruction of the alveoli and capillaries.
But, through the use of ML and DL these task of analysis can be simplified drastically.

4.3. Data Sources

This study is based on Xray images data, MERS-COV data, Pneumonia data. This data is real data which is collected from Kaggle and real data. We have chosen reliable data in particular disease: MERS-CoV , Pneumonia. The parameters of our prediction models are famous and easily available and the parameters are:

Data sources comprises of Xray images taken from various medical and government sponsored institutional websites for pulmonary diseases.
These images are filtered to match the perfect image criteria that fits the disease classification. Then X-ray images are pre-processed to standardize the input size, normalize pixel values, and enhance image quality (e.g., noise reduction, contrast adjustment).

1.Type: Chest X-ray images are the primary data source.
2.Size: A large dataset with enough labelled images for each disease category (COPD, pneumonia, MERS-CoV, and healthy controls) is crucial. We have Aim thousands of images per category to ensure model generalizability.
3.Quality: High-resolution, standardized X-ray images with consistent acquisition protocols are preferred to minimize noise and artifacts that could affect model performance.
4.Labeling: Accurate and consistent labelling of images with the corresponding disease category is essential. We have used multiple radiologists for labelling to minimize bias.

5.Data Augmentation Techniques: Techniques like rotation, flipping, and adding noise is used to artificially expand the dataset and improve model robustness to variations in real-world images.

5 PROPOSED METHOD

We have proposed a conceptual system centered around machine learning models, comprising five crucial steps: ETL (Extract, Transform, Load), feature engineering, model training, evaluation, and model deployment.
In the ETL phase, we initially collect Xray image data from various sources, followed by a series of transformations, cleaning, and processing steps specific to each data source. The refined dataset is ultimately merged and loaded into a centralized storage system.
The feature engineering phase involves the application of data analysis techniques to unveil hidden insights within the dataset, with the primary goal of preparing the data for machine learning models.
In model training we have made use of Dense net in CNN algorithm which generally consists of deep densely structured layers within the network.
Using dense net mitigates the Gradient Problem and promotes the feature reuse, increasing the precision of the model. Use of this architecture can help in extracting precise and fragile characteristics hidden in the images which can improve the result accuracy.

5.1 Dense net Architecture:

In dense net architecture each layer is connected to every other layer in feed-forward fashion.
Use of feed-forward fashion simplifies the task of reducing the gradient and make use of every feature as each layers are connected directly.
For an L-layer network, there are [L(L+1)]/2 direct connections. This ensures that the feature maps learned by each layer are used as inputs to all subsequent layers.

Mathematically speaking the output of l-th layer H' is defined as,

H'= H*([H1,H2,H3,...H"])

Here the [] brackets defines the concatenation of the inputs from all previous layers
H* is the composite function comprising of batch normalization, ReLU, convolution applied to the concatenation of feature maps from layers 0 to l-1
H" is the l-1 th layer in the network.
H1,H2,H3.. are all the output from the respective layers.

Composite function-
Each layer in the composite function consists of 3 operations which are Batch Normalization, ReLU and convolution

5.1.1.Batch Normalization-

Batch normalization is a technique generally used to standardize the inputs for each layer by adjusting and scaling the activations.
This ensures that each activation has a mean of 0 and a standard deviation that is variance of 1, this helps in speeding up the training process
and reducing the covariate shift.

5.1.2.ReLU-

Rectified Linear Unit is the most commonly used activation function in deep networks employs a Max function to the set of inputs which
introduces the non-linearity to the network by outputting the input directly as 1 if it is positive and 0 if it is not.
Mathematically it can be given as

ReLU(x)= max(0,x)

5.1.3.Convolution-

The convolutional layer is the core building block of CNNs. It applies a set of learnable filters (also known as kernels) to the input feature maps,
performing spatial convolution operations. Each filter extracts specific features from the input, such as edges, textures, or patterns,
by sliding over the input feature maps and computing element-wise multiplications followed by summations.
The output of a convolutional layer is referred to as a feature map or activation map, which represents the presence of different features within the input data.
These feature maps are then passed through activation functions (such as ReLU) to introduce non-linearity before being forwarded to subsequent layers in the network.

The composite function undertakes these three operations and can be shown as,

H*(x)= Conv(ReLU(BNN(x)))

5.2 Growth rate:

Growth rate represents the number of filters added to each layer. Filters plays an important role in extracting the characteristics and features of the images.
If k1 is the number of initial channels added to the initial layer then the total number of channels added to the l-the layer can be given as,

k + k1.(l-1)

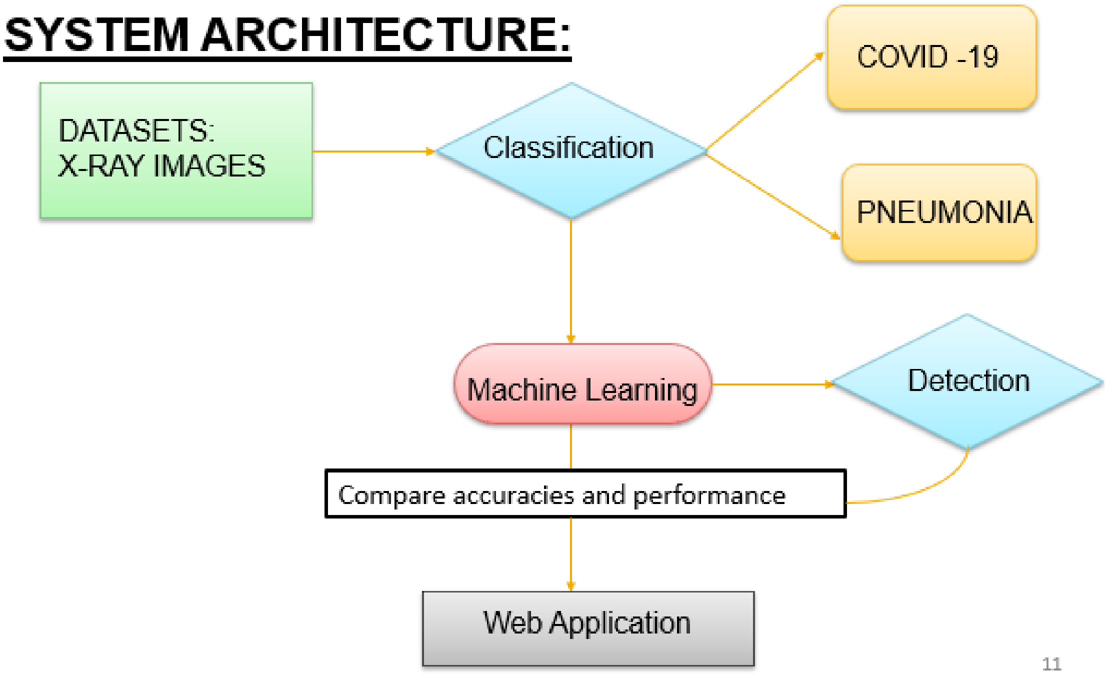
where k is the growth rate.
This linear growth rate helps manage the complexity and size of the model.

5.3 Dense Block:

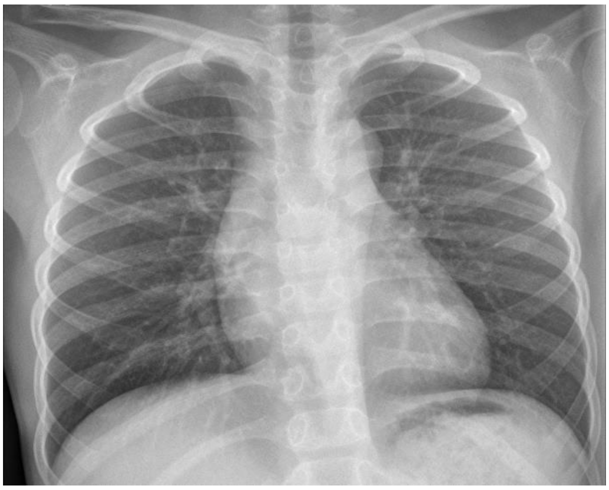
A Dense Block consists of multiple layers where each layer receives the feature maps of all preceding layers as input. If there are
n layers in a block, the output of the block is the concatenation of all these layers outputs:

Output of Dense Block=[H0,H1,…,Hn−1]

5.4 System Architecture



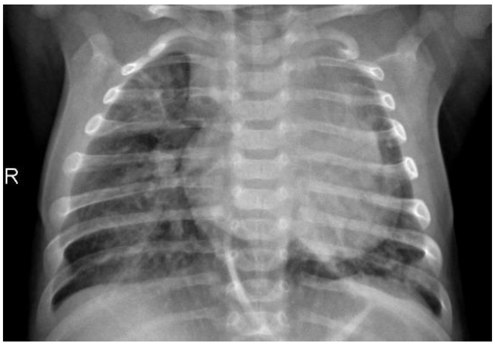
6 RESULTS & DISCUSSIONS



6.1 Normal Xray of Patient

Result:

6.2 Covid-19 Detected Xray



6.3 Pneumonia Detected Xray

We have taken the dataset of normal images, Covid 19 Xray images , pneumonia images which have helped us to classify and detect the COPD with the help of machine learning models.

7 CONCLUSION & FUTURE SCOPE

In this study we proposed a new method to detect lung diseases, like Covid-19 and pneumonia, using chest X-ray images. This method uses different techniques including image improvement, focusing on specific lung areas, and extracting key features. To improve accuracy while using less computing power, a deep learning model is introduced. This suggests promise for its future use. Improvements for the future could include making the model adaptable to different datasets and able to analyze disease severity.

These findings will assist doctors in choosing suitable models for various image analysis methods, which will be important when time and resources are limited in a pandemic scenario like the present COVID-19. As future work, the proposed method could be implemented on a dataset with more classes of pulmonary diseases such as asthma, chronic obstructive pulmonary disease, pulmonary fibrosis, pneumonia, lung cancer and COVID-19.

8 REFERENCE

Golabbakhsh, M., Ramezankhani, A., Aghamohamadi, A., & Sharifi, A. 134(34) (2020). A review of machine le arning methods for the early diagnosis of chronic obstructive pulmonary disease. Journal of Ambient Intelligence and Humanized Computing.

Aggarwal, R., & Soundrapandian, S.629(57) (2018). A review on automated COPD detection using deep learning techniques. Journal of King Saud University-Computer and Information Sciences.

Saha, S., Shit, S., & Balasubramanian, V.450(34) (2021). Deep learning and transfer learning based automated prediction of chronic obstructive pulmonary disease. Biocybernetics and Biomedical Engineering, Lee, Y. C., Hung, S. H., Su, M. C., & Chou, P. C. 589(30)(2019). A hybrid approach of feature selection and machine learning for COPD disease classification.

Bhagat, M., Syed, K., Singh, S., et al. 789(67)(2019). Predicting COPD readmissions using ensemble machine learning. Proceedings of the IEEE International Conference on Healthcare Informatics.

Chen, C. H., & Huang, Y. S. 344(45)(2013). Improving the accuracy of disease prediction in computer-assisted diagnosis by data mining for clinical and laboratory data. Journal of Medical Systems.

Pérez, J. M., & Núñez, L. A.567(76) (2008). Comparison of predictive models for the forced expiratory volume in one second as outcome in COPD.